Big Data and Computing Visions

www.bidacv.com

Big. Data. Comp. Vis. Vol. 2, No. 1 (2022) 9–17.

Paper Type: Original Article

A Comprehensive Review of Big Data Applications

Soheil Fakheri *

[1] Department of Computer Engineering, Ayandegan Institute of Higher Education, Tonekabon, Iran; fakheri.soheil@gmail.com.

Citation:

Fakheri, S. (2022). A comprehensive review of big data applications. *Big data and computing visions, 2*(1), 9-17.

Received: 18/01/2022 Reviewed: 09/02/2022 Revised: 10/03/2022 Accepted: 19/03/2022

Abstract

Analysis of big data has been presented as an advanced analytical technology involving large-scale and complex applications. In this paper, we review the general background of big data, and focus on data generation and data analysis. Then, we examine the several representative applications of big data, including enterprise management, Internet of Things, online social networks. These discussions aim to provide a comprehensive overview to readers of this exciting area.

Keywords: Big data application, Data analysis, Text analysis, Web analysis, Multimedia analysis.

1 | Introduction

Over the past 20 years, large-scale data has increased in various fields. Under the explosive rise of global data, the term big data is mainly used to describe massive data sets. Compared to traditional data sets, big data typically contains piles of unstructured data that require more real-time analysis. In addition, big data also brings new opportunities to discover new values, helps us gain a deep understanding of hidden values, and also incurs new challenges, as usual, how to effectively organize and manage such datasets [1].

Currently although the importance of big data is generally recognized, people still have different opinions about defining it. In general, big data should mean data sets that cannot be understood, acquired, managed, and processed by traditional IT and hardware software/tools at a tolerable time. In 2010, Apache Hadoop defined big data as "data sets that could not be recorded, managed and processed by public computers within an acceptable range." According to this definition, in May 2011, McKinsey & Company, a global consulting agency, announced big data as the next frontier of innovation, competition, and productivity [2].

Corresponding Author: fakheri.soheil@gmail.com

http://dx.doi.org/10.22105/bdcv.2022.325256.1041

Fakheri| Big. Data. Comp. Vis. 2(1) (2022) 9-17

Big data should mean such data sets that cannot be obtained, stored, and managed by classic database software. This definition consists of two sentences: first, that the volume of data sets that meet the big data standard is changing, and may grow over time or with technological advances; second, the volume of data sets that meet the big data standard in different applications is different. Currently, big data generally spans from multiple TB to multiple PB [2].

Today, big data related to the service of internet companies is growing rapidly. For example, Google processes data from hundreds of petabites (PB), Facebook generates log data of more than 10 PB per month, Baidu, a Chinese company, processes data from dozens of PB's, and Taobao, a subset of Alibaba, generates data from dozens of terabits (TB) for online transactions per day. As internet services developed, queryed indicators and contents were growing rapidly. So search engine companies had to face the challenges of handling such big data. Google created the GFS [3] and MapReduce [4] programming models to tackle the challenges brought by data management and internet-scale analytics. In addition, content produced by users, sensors, and other ubiquitous data sources also displayed onerous data streams that required a fundamental change to the computing architecture and large-scale data processing mechanism. In January 2007, database software pioneer Jim Gray called such a development "Paradigm IV" [5]. He also thought the only way to deal with such a paradigm was to develop a new generation of computational tools to manage, visualize and analyze massive data. In June 2011, another milestone event occurred; EMC/IDC published a research report titled Extracting Values from Chaos [6], which introduced the concept and potential of big data for the first time.

In the rest of paper, we first review the evolution of data sources. In Section 2 we examine six of the most important areas of data analysis including organized data analysis, text analysis, website analysis, multimedia analytics, network analysis, and mobile analytics. Finally, conclusion of paper provided in Section 3.

2 | Big Data Application

Big data analytics can provide useful values through judgment, suggestions, support, or decision making. However, data analysis involves a wide range of applications, which often change and are very complex.

2.1 | Key Applications of Big Data

Evolution of the program recently, big data analytics has been proposed as an advanced analytical technology, which typically involves large-scale and complex applications under certain analytical methods. In fact, data driven applications have emerged in the past decades. By the early 1990s, for example, BI had become a dominant technology for business applications, and network search engines emerged based on massive data mining processing in the early 21st century. Some potential and influential applications from different backgrounds and their data and analysis features are discussed as follows.

2.1.1 | Evolution of business applications

The first business data generally structured data, which was collected by companies from legacy systems and then stored in RDBMSs. The analytical techniques used in such dominant systems were in the 1990s and were intuitive and simple, for example, in reporting forms, dashboards, querying with terms, search-based business intelligence, online transaction processing, interactive visualization, scorecards, predictive modeling, and data mining [7]. Since the beginning of the 21st century, Global Wide Web Networks and Networks (WWW) provide a unique opportunity for organizations to have online displays and interact directly with customers. Abundant products and customer information such as clickstream data logs and user behavior can be obtained from WWW. Product layout optimization, customer trade analysis, product offers, and market structure analysis can be done by text analysis and website mining techniques. The quantity of mobile phones and tablets surpassed laptops and PCs for the first time in

2011. Mobile phones and the Internet of Things based sensors are opening up a new generation of innovation applications, and require significantly larger capacity to support location sensing, people-oriented, and field-conscious practice.

2.1.2 | Evolution of network applications

The early generation of the Internet mainly provided email and WWW services. Text analysis, data mining, and web page analysis have been applied to extract email contents and building search engines. Today, most applications are web-based regardless of their field goals and design. Network data account for a major percentage of global data volume. The web has become a common platform for interconnected pages, full of different types of data, such as text, images, audio, videos, and interactive content, etc. Therefore, a large amount of advanced technologies used for semi-structured or unstructured data emerged at the right moment. For example, image analysis can extract useful information from images, (e.g., facial recognition). Multimedia analytics technologies can be applied to automated video surveillance systems for businesses, law enforcement, and military applications. Since 2004, online social media, such as internet forums, online communities, blogs, social networking services, and multimedia websites, offer users plenty of opportunities to create, upload and share content [8].

2.1.3 | Evolution of scientific application

Scientific research in many fields is gaining massive data with highthroughput sensors and tools such as astrophysics, oceanography, genomics, and environmental research. The American National Science Foundation recently announced the BIGDATA program to promote efforts to extract knowledge and insight from large and complex digital data collections. Some scientific research disciplines have developed big data platforms and have achieved useful results. In biology, for example, iPlant [9] applies network infrastructure, physical computing resources, coordination environment, virtual machine resources, inter-operational analysis software, and data services to help researchers, educators, and students enrich plant sciences. iPlant datasets have high types in form, including specifications or reference data, experimental data, analog or model data, observation data, and other derivative data.

2.2 | Data Analysis Models

We can divide data analytics research into six key technical areas, as one, structured data analysis, text data analysis, web data analysis, multimedia data analysis, network data analysis, and mobile data analysis. Such a classification aims to emphasize the characteristics of the data, but some areas may use similar basic technologies. Since data analysis has wide scope and comprehensive coverage is not easy, we will focus on key problems and technologies in data analysis in the following discussions.

2.2.1 | Structured data analysis

The structure of data analysis business applications and scientific research may produce massive structured data, from which management and analysis rely on mature commercial technologies, such as RDBMS, data warehouse, OLAP, and BPM (Business Process Management) [10]. Data analysis is mainly based on data mining and statistical analysis, both of which have been well studied over the past 30 years. However, data analysis is still a very active field of research and new functional demands will lead to the development of new methods. For example, statistical machine learning based on accurate mathematical models and powerful algorithms has been applied to anomaly detection [11] and energy control [12]. Exploiting the characteristics of data, time and space mining can extract knowledge structures hidden in high-speed data streams and sensors [13]. Driven by privacy protections in e-commerce, e-government, and healthcare applications, data mining privacy protection is an emerging research field [14]. Over the past decade, process extraction has become a new research field especially in process analysis with event data [15].

2.2.2 | Text data analysis

Text data analysis is the most common data storage format of text, as such, emails, business documents, web pages, and social media. Therefore, text analysis is deemed to have more business-based potential characteristics than structured data. In general, text analysis is a process for extracting useful information and knowledge from unstructured text. Text mining is interdisciplinary, which includes data retrieval, machine learning, statistics, computational linguistics, and data mining in particular. Most text extraction systems are based on text phrases and Natural Language Processing (NLP), with a greater emphasis on the latter. NLP allows computers to analyze, interpret and even generate text. Some common NLP methods include lexical acquisition, word sense disambiguation, part-speech labeling, and free grammar of possibilities context [16]. Some NLP-based techniques have been applied to text extraction, including data extraction, subject models, text summarization, classification, clustering, question answers, and comment mining.

2.2.3 | Web data analytics

Web data analytics data analysis has emerged as an active research field. It aims to automatically retrieve, extract, and evaluate information from web documents and services so as to discover useful knowledge. Web analytics relates to several research areas including databases, data recovery, NLP, and text mining. According to the different sections extracted, we categorize web data analysis into three related areas: extracting web content, extracting web structure, and extracting web usage [17]. Extracting web content is a process of discovering useful knowledge on web pages, which generally includes a variety of data, such as text, image, audio, video, code, metadata, and hyperlink.

2.2.4 | Multimedia analysis

Research on image, audio, and video extraction has just been called multimedia analysis, which will be discussed in Section 6.1.5. Since most of the web content data is unstructured text data, research on web data analysis mainly centers around text and hypertext. Text mining is discussed in Section 6.1.3, while hypertext extraction involves extracting semi-structured HTML files containing hyperlink. Supervised learning and classification plays important roles in hyperlink extraction, such as, email, newsgroup management, and web catalog maintenance [18]. Extracting web content can be done by two methods of data retrieval and database method. Data recovery mainly helps or improves data reviews, or filters user information according to deductions or configuration documents. The purpose of the database method is to simulate and integrate data into the web, to conduct more complex queries than searches based on keywords. Web structure extraction includes models for discovering web link structures. Here the structure refers to related schematic charts on a website or among multiple websites. Models are built on topological structures provided with hyperlinks with or without link descriptions. Such models reveal similarities and correlations among different websites and are used to categorize website pages. Pagerank [19] and Clever [20] make full use of the model to look up the relevant website pages. Thematic reptile is another successful case using the model [21]. Extracting web use is aimed at mining auxiliary data generated by web dialogs or activities. Web mining content and web mining structure use basic web data. Web usage data includes access logs on web servers and proxy servers, browser history records, user profiles, registration data, user meetings or transactions, cache, user queries, bookmarking data, mouse clicks and scrolls, and any other types of data generated through web interaction.

As web services and Web2.0 are becoming mature and popular, web usage data will increasingly have a high variety. Web use extraction plays key roles in personalized space, e-commerce, network privacy/security, and other emerging areas. For example, participatory recommendation systems can personalize e-commerce using different user preferences.

Multimedia data analysis of multimedia data (mainly including images, audio, and videos) have been growing at an amazing pace, which extracted useful knowledge and understood semantemes with analysis.

Because multimedia data is intigent and most such data contains richer information than plain structured data or textual data, data extraction faces a huge challenge of semantic differences. Research on multimedia analytics covers many disciplines. Some recent research priorities include multimedia summarization, multimedia description, multimedia index and recovery, multimedia offer, and multimedia event recognition, etc. Audio summary can be done by extracting highlight words or phrases from metadata or synthesizing a new representation. Video summarization is to interpret the most important or representative video content sequence, and can be static or dynamic. Static video summarization methods use a key frame sequence or context-sensitive key frames to represent a video. Such methods are simple and have been applied to many business applications (such as Yahoo, Alta Vista and Google), but their performance is poor. Dynamic summarization methods use a series of video frames to show a video, taking other smooth actions to make the final summary look more natural. The authors of [22], propose a subject-oriented multimedia summarization system (TOMS) that can automatically summarize important information in a video belonging to a particular subject area, based on a specific set of features extracted from the video. A multimedia explanation inserts tags to describe the contents of images and videos at both syntax and semantic levels. With such tags, it is easy to manage, summarize, and retrieve multimedia data.

Since manual annotation is both time and labor intensive, automated annotation without any human interventions becomes awesome. The main challenge for automatic multimedia description is semantic difference. Although much progress has been made, the performance of existing automated writing methods still needs improvement. A lot of efforts are currently underway to simultaneously discover both manual and automated multimedia mentions [23]. Multimedia indexing and retrieval includes describing, storing, and organizing multimedia information and helping users easily and quickly look at multimedia resources [24]. In general, multimedia indexing and retrieval consists of five procedures: structural analysis, feature extraction, data mining, classification and indexing, querying and retrieval [25]. Structural analysis aims to segment a video into several semantic structural elements, including lens boundary detection, key frame extraction, and scene separation, etc. According to the result of structural analysis, the second method is feature extraction which mainly involves extracting most of the features of key frames, objects, texts, and gestures that are the foundation of indexing and video retrieval. Data mining, classification, and references to the use of extracted features to find video content modes and put videos into scheduled categories so as to generate video indicators. After receiving the query, the system will use a similarity measurement method to look at a candidate video. Optimizes the recovery result of relevant feedback. The multimedia recommendation is to recommend specific multimedia content according to users' preferences. It proves to be an effective approach to providing personalized services. Most existing recommendation systems can be categorized in content-based systems and collaboration-based policing systems. Content-based methods identify the overall characteristics of users or their interesting, and recommend users for other content with similar features. These methods rely largely on measuring content similarity, but most of them are troubled by limitations of analysis and excessive specifications.

Collaborative filtering methods identify groups with similar interests and recommend content for group members according to their behavior [26]. Now, a mixed method is introduced, which integrates the benefits of the two types of above methods to improve the quality of the recommendation [27]. The U.S. National Institute of Standards and Technology (NIST) launched the TREC Video Recovery Assessment to detect an event in video clips based on Event Kit, which contains some text descriptions of concepts and video samples [28] and [29], the author proposed a new algorithm for special multimedia event recognition using several positive training examples. Research on video event detection is still in its infancy, focusing primarily on sporting or news events, performances or abnormal events in video monitoring, and other similar events with repetitive patterns.

Fakheri| Big. Data. Comp. Vis. 2(1) (2022) 9-17

2.2.5 | Network data analysis

Network data analysis has evolved from initial quantitative analysis [30] and sociological network analysis [31] to emerging online social network analysis at the beginning of the 21st century. Many online social networking services, including Twitter, Facebook, and LinkedIn, etc. have become increasingly popular over the years. Such online social networking services generally include massive relevant data and content data. Link data is mainly in the form of graphical structures describing communications between the two entities. Content data includes text, image, and other network multimedia data. Rich content on such networks brings both unprecedented challenges of nd opportunities for data analysis. According to the data-driven perspective, existing research in the fields of social networking services can be classified into two categories: link-based structural analysis and content-based analysis [32]. Research on link-based structural analysis has always been conducted on link prediction, community discovery, social network development, and social influence analysis, etc. SNS may be visualized as a graph, in which each vertex matches a user and the edges match the correlations among users. Since SNS are dynamic networks, new vertices and edges are constantly added to graphs. Link prediction predicts the possibility of future connection between two vertices. Many techniques can be used to predict grafting, for example, feature-based classification, possibility methods, and linear algebra. Feature-based classification is to select a group of features for a vertex and use the existing link information to generate binary classifiers to predict the future link [33].

Prometheories methods aimed at constructing models for connection possibilities among apexes in SNS [34]. Linear algebra calculates the similarity between the two vertices according to the same singular matrix [35]. A society is represented by a subgraph matrix, in which the binding edges of the vertices have a high density under the graph, while the edges between the two subgraphs have a much lower density [36]. Many methods have been proposed and studied for community diagnosis, most of which are target functions based on tapology based on the concept of capturing the structure of society. Do et al. use the property overlapping communities in real life to propose an effective large-scale SNS community recognition method [37]. The purpose of research on SNS is to look for a rule and deduction model to interpret the evolution of the network. Some empirical studies have found that proximity bias, geographical constraints, and other factors play important roles in the evolution of SNS [38]-[40], and some generational methods are suggested to help design the network and system [41]. Social influence refers to a case where people under the influence of others change their behavior. The power of social influence depends on the relationship between individuals, network distances, time effect, and the characteristics of networks and individuals, etc. Marketing, advertising, recommendations, and other uses can benefit from social impact by quantitatively measuring the impact of individuals on others [42], [43] in general, if the duplication of content in SNS is taken into account, the performance of link-based structural analysis may be further improved. Content-based analytics on SNS is also known as social media analytics. Social media includes text, multimedia, positioning, and comments. Yet social media analysis faces unprecedented challenges. First, massive and constantly growing social media data should be automatically analyzed in a reasonable time window. Second, social media data contains a lot of noise. For example, bloggers include a lot of spam blogs, and it has inconseance tweets on Twitter as well. Third, SNS dynamic networks, which often and quickly vary and are updated. Existing research on social media analysis is still in its infancy. Given that SNS contains massive information, learning to transmit in incognomous networks with the aim of transmitting knowledge information among different media [44].

2.2.6 | Mobile data analysis

Analyzing mobile data as of April 2013, Android apps provided more than 650,000 applications, covering almost all categories. By the end of 2012, monthly mobile data flows had reached 885 PB [45]. Massive data and plentiful applications demand mobile analysis, but also a few challenges. As a whole, mobile data has unique features, as such, mobile sensing, animated flexibility, noise, and plenty of redundancy. Recently, new research has been launched in the field of mobile analysis in various fields.

Since the investigation into mobile analytics has just begun, we have only introduced some recent and representative analytics applications in this section. As the number of mobile users grows and performance improves, mobile is now useful for building and maintaining communities, such as communities with geographic locations and communities based on different cultural backgrounds and interests (for example, the latest Webchat). Traditional networking communities or SNS communities are short of online interaction between members, and communities are only active when members sit against computers. In the photo, mobile phones can support rich engagement anytime, anywhere. Mobile communities are defined as a group of people with the same hobbies (as one, health, safety, and entertainment, etc.) gather on networks, meet to create a common goal, make decisions through counseling to achieve their goal, and begin to implement their plan [46]. In [47], the authors proposed a qualitative model of a mobile community. It is now widely believed that mobile community applications will greatly promote the development of the mobile industry. Recently, advances in wireless sensor, mobile communications technology, and streaming processing enable people to build a body area network to promptly monitor people's health. In general, medical data from different sensors have different characteristics in terms of characteristics, temporal and spatial relationships as well as physiological relationships, etc.

In addition, such datasets include privacy and safety protection. In [48], Garg and his colleagues introduce a multimodal transport analysis mechanism for prompt monitoring of health. In situations where only very comprehensive health-related features are available, Park and her colleagues in examined better utilization approaches. Researchers from Gjwik University College in Norway and the Biometric Drawi collaborated to develop a smartphone app that analyzes speeds while people walk and uses speed information to unlock the immune system [49]. Meanwhile, Robert Delano and Brian Pryce of the Georgia Institute of Technology developed a program called iTrem that monitors human body tremors with internal seismicity on mobile phones to cope with Parkinson's and other nervous system diseases.

3 | Conclusion

The big data technology is still in its infancy. Many key technical problems, such as cloud computing, grid computing, stream computing, parallel computing, big data architecture, big data model, and software systems supporting big data, etc. should be fully investigated. In this paper, we review the background and state-of-the-art of big data, then we focus on the some of the most important areas of data analysis including organized data analysis, text analysis, website analysis, multimedia analytics, network analysis, and mobile analytics. Finally, we introduce several key areas of the program from big data.

References

[1] Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. *2013 international conference on collaboration technologies and systems (CTS)* (pp. 42-47). IEEE. DOI: 10.1109/CTS.2013.6567202

[2] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: the next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

[3] Ghemawat, S., Gobioff, H., & Leung, S. T. (2003, October). The Google file system. *Proceedings of the nineteenth ACM symposium on operating systems principles* (pp. 29-43). https://doi.org/10.1145/945445.945450

[4] Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107-113. https://doi.org/10.1145/1327452.1327492

[5] Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). Redmond, WA: Microsoft research.

[6] Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iview*, *1142*(2011), 1-12.

[7] Sallam, R. L., Richardson, J., Hagerty, J., & Hostmann, B. (2011). Magic quadrant for business intelligence platforms. Retrieved from Gartner. http://in2015.in1.com.br/sites/default/files/magic_quadrant_for_business__210036.pdf

[8] Chen, M., Mao, S., & Liu, Y. (2014). Big data: a survey. *Mobile networks and applications*, *19*(2), 171-209. https://doi.org/10.1007/s11036-013-0489-0

[9] Goff, S. A., Vaughn, M., McKay, S., Lyons, E., Stapleton, A. E., Gessler, D., ... & Stanzione, D. (2011). The iPlant collaborative: cyberinfrastructure for plant biology. *Frontiers in plant science*, *2*, 34. https://doi.org/10.3389/fpls.2011.00034

[10] Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB endowment*, *5*(12), 2032-2033.

[11] Baah, G. K., Gray, A., & Harrold, M. J. (2006, November). On-line anomaly detection of deployed software: a statistical machine learning approach. *Proceedings of the 3rd international workshop on Software quality assurance* (pp. 70-77). https://doi.org/10.1145/1188895.1188911

[12] Moeng, M., & Melhem, R. (2010, May). Applying statistical machine learning to multicore voltage & frequency scaling. *Proceedings of the 7th ACM international conference on computing frontiers* (pp. 277-286). https://doi.org/10.1145/1787275.1787336

[13] Gaber, M. M., Zaslavsky, A., & Krishnaswamy, S. (2005). Mining data streams: a review. *ACM sigmod record*, *34*(2), 18-26. https://doi.org/10.1145/1083784.1083789

[14] Verykios, V. S., Bertino, E., Fovino, I. N., Provenza, L. P., Saygin, Y., & Theodoridis, Y. (2004). State-of-the-art in privacy preserving data mining. *ACM sigmod record*, *33*(1), 50-57. https://doi.org/10.1145/974121.974131

[15] Van Der Aalst, W. (2012). Process mining: overview and opportunities. *ACM transactions on management information systems (TMIS)*, *3*(2), 1-17. https://doi.org/10.1145/2229156.2229157

[16] Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

[17] Pal, S. K., Talwar, V., & Mitra, P. (2002). Web mining in soft computing framework: relevance, state of the art and future directions. *IEEE transactions on neural networks*, *13*(5), 1163-1177. DOI: 10.1109/TNN.2002.1031947

[18] Chakrabarti, S. (2000). Data mining for hypertext: a tutorial survey. *ACM SIGKDD explorations newsletter*, *1*(2), 1-11. https://doi.org/10.1145/846183.846187

[19] Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, *30*(1-7), 107-117. https://doi.org/10.1016/S0169-7552(98)00110-X

[20] Konopnicki, D., & Shmueli, O. (1995, September). W3qs: a query system for the world-wide web. Proceeding of the 21th VLDB Conference (Vol. 95, pp. 54-65). Zurich, Switzerland. http://www.vldb.org/conf/1995/P054.PDF

[21] Chakrabarti, S., Van den Berg, M., & Dom, B. (1999). Focused crawling: a new approach to topic-specific Web resource discovery. *Computer networks*, *31*(11-16), 1623-1640. https://doi.org/10.1016/S1389-1286(99)00052-3

[22] Ding, D., Metze, F., Rawat, S., Schulam, P. F., Burger, S., Younessian, E., ... & Hauptmann, A. (2012, June). Beyond audio and video retrieval: towards multimedia summarization. *Proceedings of the 2nd ACM international conference on multimedia retrieval* (pp. 1-8). https://doi.org/10.1145/2324796.2324799

[23] Wang, M., Ni, B., Hua, X. S., & Chua, T. S. (2012). Assistive tagging: a survey of multimedia tagging with human-computer joint exploration. *ACM computing surveys (CSUR)*, *44*(4), 1-24. https://doi.org/10.1145/2333112.2333120

[24] Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: state of the art and challenges. *ACM transactions on multimedia computing, communications, and applications (TOMM)*, *2*(1), 1-19. https://doi.org/10.1145/1126004.1126005

[25] Hu, W., Xie, N., Li, L., Zeng, X., & Maybank, S. (2011). A survey on visual content-based video indexing and retrieval. *IEEE transactions on systems, man, and cybernetics, part C (applications and reviews)*, *41*(6), 797-819. DOI: 10.1109/TSMCC.2011.2109710

[26] Park, Y. J., & Chang, K. N. (2009). Individual and group behavior-based customer profile model for personalized product recommendation. *Expert systems with applications*, *36*(2), 1932-1939. https://doi.org/10.1016/j.eswa.2007.12.034

[27] Barragáns-Martínez, A. B., Costa-Montenegro, E., Burguillo, J. C., Rey-López, M., Mikic-Fonte, F. A., & Peleteiro, A. (2010). A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. *Information sciences*, *180*(22), 4290-4311. https://doi.org/10.1016/j.ins.2010.07.024

[28] Naphade, M., Smith, J. R., Tesic, J., Chang, S. F., Hsu, W., Kennedy, L., ... & Curtis, J. (2006). Large-scale concept ontology for multimedia. *IEEE multimedia*, *13*(3), 86-91. DOI: 10.1109/MMUL.2006.63

Fakheri| Big. Data. Comp. Vis. 2(1) (2022) 9-17

[29] Ma, Z., Yang, Y., Cai, Y., Sebe, N., & Hauptmann, A. G. (2012, October). Knowledge adaptation for ad hoc multimedia event detection with few exemplars. *Proceedings of the 20th ACM international conference on Multimedia* (pp. 469-478). https://doi.org/10.1145/2393347.2393414

[30] Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the national academy of sciences*, *102*(46), 16569-16572. https://doi.org/10.1073/pnas.0507655102

[31] Bower, D. F. (2005). Six degrees: the science of a connected age. *Complicity: an international journal of complexity and education*, *2*(1). DOI: https://doi.org/10.29173/cmplct8734

[32] Aggarwal, C. C. (2011). An introduction to social network data analytics. In *Social network data analytics* (pp. 1-15). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-8462-3_1

[33] Scellato, S., Noulas, A., & Mascolo, C. (2011, August). Exploiting place features in link prediction on location-based social networks. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1046-1054). https://doi.org/10.1145/2020408.2020575

[34] Ninagawa, A., & Eguchi, K. (2010, March). Link prediction using probabilistic group models of network structure. *Proceedings of the 2010 ACM symposium on applied computing* (pp. 1115-1116). https://doi.org/10.1145/1774088.1774323

[35] Dunlavy, D. M., Kolda, T. G., & Acar, E. (2011). Temporal link prediction using matrix and tensor factorizations. *ACM transactions on knowledge discovery from data (TKDD)*, *5*(2), 1-27. https://doi.org/10.1145/1921632.1921636

[36] Leskovec, J., Lang, K. J., & Mahoney, M. (2010, April). Empirical comparison of algorithms for network community detection. *Proceedings of the 19th international conference on world wide web* (pp. 631-640).

[37] Du, N., Wu, B., Pei, X., Wang, B., & Xu, L. (2007, August). Community detection in large-scale social networks. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on web mining and social network analysis* (pp. 16-25). https://doi.org/10.1145/1348549.1348552

[38] Garg, S., Gupta, T., Carlsson, N., & Mahanti, A. (2009, November). Evolution of an online social aggregation network: an empirical study. *Proceedings of the 9th ACM SIGCOMM conference on internet measurement* (pp. 315-321). https://doi.org/10.1145/1644893.1644931

[39] Allamanis, M., Scellato, S., & Mascolo, C. (2012, November). Evolution of a location-based online social network: analysis and models. *Proceedings of the 2012 internet measurement conference* (pp. 145-158). https://doi.org/10.1145/2398776.2398793

[40] Gong, N. Z., Xu, W., Huang, L., Mittal, P., Stefanov, E., Sekar, V., & Song, D. (2012, November). Evolution of social-attribute networks: measurements, modeling, and implications using google+. *Proceedings of the 2012 internet measurement conference* (pp. 131-144). https://doi.org/10.1145/2398776.2398792

[41] Zheleva, E., Sharara, H., & Getoor, L. (2009, June). Co-evolution of social and affiliation networks. *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1007-1016). https://doi.org/10.1145/1557019.1557128

[42] Tang, J., Sun, J., Wang, C., & Yang, Z. (2009, June). Social influence analysis in large-scale networks. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 807-816). https://doi.org/10.1145/1557019.1557108

[43] Li, Y., Chen, W., Wang, Y., & Zhang, Z. L. (2013, February). Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. *Proceedings of the sixth ACM international conference on web search and data mining* (pp. 657-666). https://doi.org/10.1145/2433396.2433478

[44] Dai, W., Chen, Y., Xue, G. R., Yang, Q., & Yu, Y. (2008). Translated learning: transfer learning across different feature spaces. *Advances in neural information processing systems*, *21*, 353-360.

[45] Pepper, R. (2013). *Cisco visual networking index (VNI) global mobile data traffic forecast update*. Retrieved from Cisco. https://www.gsm.org/spectrum/wp-content/uploads/2013/03/Cisco_VNI-global-mobile-data-traffic-forecast-update.pdf

[46] Rhee, Y., & Lee, J. (2009). On modeling a model of mobile community: designing user interfaces to support group interaction. *Interactions*, *16*(6), 46-51. https://doi.org/10.1145/1620693.1620705

[47] Han, J., Li, Z., & Tang, L. A. (2010, April). Mining moving object, trajectory and traffic data. *International conference on database systems for advanced applications* (pp. 485-486). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-12098-5_56

[48] Garg, M. K., Kim, D. J., Turaga, D. S., & Prabhakaran, B. (2010, March). Multimodal analysis of body sensor network data streams for real-time healthcare. *Proceedings of the international conference on multimedia information retrieval* (pp. 469-478). https://doi.org/10.1145/1743384.1743467

[49] Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.